

DOI: 10.13382/j.jemi.B2306912

基于文本图像修正的两阶段船名识别框架*

卢鹏涛^{1,2} 蒋雯³ 黄菊^{1,2} 孙水发⁴ 汪方毅^{1,2}

(1. 湖北省水电工程智能视觉监测重点实验室 宜昌 443002; 2. 三峡大学计算机与信息学院 宜昌 443002;
3. 长江宜昌通信管理局 宜昌 443001; 4. 杭州师范大学 杭州 311121)

摘要: 船舶名称(牌照)识别在水路运输系统中发挥着重要的作用。针对船舶名称在内河航道中目标较小且航道两岸观测船舶存在较大倾斜角度导致难以识别的问题, 提出一个以自然场景文本检测算法(differentiable binarization, DB)和文本识别算法(convolutional recurrent neural network, CRNN)为基础的船舶名称自动识别框架(automatic ship name identification, ASNI), ASNI包括以下3个部分: 船名检测、文本图像修正和识别, 其中, 船名文本图像修正由船名矫正模块和超分辨率重建模块构成。首先, 该框架利用DB算法对图像船名候补区域特征进行自适应尺度融合处理获取特征图, 通过特征映射预测生成的二值图像寻找连接区域, 以此获得船名感兴趣区域(ROI)。其次, 在船名检测之后引入船名矫正模块, 基于透视变换对ROI中船名不规则文本进行矫正。此外, 设计超分辨率重建模块, 对矫正后的船名图像进行超分辨率重建处理, 以提高船名图像的分辨率。最后, 利用CRNN算法对文本图像修正后的ROI中船名进行识别得到最终结果。通过在内河航道船舶数据集(ship license plate, SLP)上进行训练和测试, 最终实验结果显示, ASNI框架对船舶识别的平均准确率为87.50%, 相比于基础框架提升了3.12%。本文设计的框架有效解决了因分辨率不足和倾斜导致船舶识别不准确的问题, 相比基础框架, ASNI有更好的识别效果。

关键词: 船舶名称识别; 文本检测; 透视变换; 超分辨率

中图分类号: TN911.73 **文献标识码:** A **国家标准学科分类代码:** 520.60

Two-stage vessel name recognition framework based on text image correction

Lu Pengtao^{1,2} Jiang Wen³ Huang Ju^{1,2} Sun Shuifa⁴ Wang Fangyi^{1,2}

(1. Hubei Key Laboratory of Intelligent Vision Based Monitoring for Hydroelectric Engineering, China Three Gorges University, Yichang 443002, China; 2. College of Computer and Information Technology, China Three Gorges University, Yichang 443002, China; 3. Yangtze Yichang Communications Authority, Yichang 443001, China; 4. Hangzhou Normal University, Hangzhou 311121, China)

Abstract: The recognition of vessel names (license plates) plays a crucial role in waterway transportation systems. Addressing the challenge of identifying vessel names in inland waterways, where targets are relatively small and vessels are observed at significant angular inclinations on both sides of the waterway, we propose an automatic ship name identification (ASNI) framework based on the differentiable binarization (DB) natural scene text detection algorithm and the convolutional recurrent neural network (CRNN) text recognition algorithm. ASNI comprises three main components: ship name detection, text image correction, and recognition. The text image correction component consists of a ship name correction module and a super-resolution reconstruction module. Firstly, the framework utilizes the DB algorithm to perform adaptive scale fusion processing on the candidate regions of vessel names in images, generating feature maps. Feature mapping is used to predict and generate binary images to identify connected regions, thereby obtaining regions of interest (ROI) containing vessel names. Subsequently, after ship name detection, a ship name correction module is introduced to rectify irregular text within the ROI using perspective transformation. Furthermore, a super-resolution reconstruction module is designed to enhance the resolution of the corrected vessel name images. Finally, the CRNN algorithm is employed to recognize vessel names within the corrected text images in the ROI, yielding the ultimate results. Through training and testing on the ship license

收稿日期: 2023-09-20 Received Date: 2023-09-20

* 基金项目: 国家自然科学基金(61871258)项目资助

plate (SLP) dataset specific to inland waterways, experimental results demonstrate that the ASNI framework achieves an average accuracy of 87.50% in vessel recognition, representing a 3.12% improvement over the baseline framework. The framework presented in this paper effectively addresses issues related to low resolution and angular inclinations leading to inaccurate vessel recognition. Compared to the baseline framework, ASNI exhibits superior recognition performance.

Keywords: vessel name recognition; text detection; perspective transform; super-resolution

0 引言

随着我国经济的快速发展,船舶数量也随着日益增长的水路运输需求而增长,港口和航道也变得越来越拥挤。船名作为标识船舶身份的重要信息之一,对船名的识别尤为重要,准确高效的船舶名称识别方案对水路航运的管理具有重要意义。

传统的船舶识别依靠船舶自动识别系统(automatic identification system, AIS),该系统能自动交换船位、航速、航向、船名等重要信息,船舶的自动识别依靠 AIS 在其高频覆盖范围内发送这些重要信息和接收其他船舶的信息来实现^[1]。但是 AIS 装载在船舶上,与岸上的交流属于被动应答,且船上的 AIS 可以人为关闭,有些小型船舶甚至没有装载 AIS,存在非常大的局限性和隐患,所以主动对船舶牌照进行识别是非常有必要的。

近年来,随着计算机技术的不断发展,应用深度学习技术对自然场景下文本进行检测及识别成为备受关注的研究领域^[2-4]。自然场景中的文本可能出现在各种复杂背景中,同时场景文本在颜色、字体和空间布局等方面变化显著,这使得文本的定位和提取变得更加困难。而车牌识别^[5-10]和船舶牌照识别^[11-13]作为该领域下的研究热点,其面临的挑战更具独特性。在面对复杂背景的情况下,汽车牌照识别通常需要首先消除潜在的噪声干扰,如路标和广告等元素,以确保有效的定位和最终的识别结果。与汽车识别不同的是,船舶通常出现在河流和海洋中,并且大多数情况下只有船体本身包含文本,一般为船舶名称。因此,可以将船舶名称定位问题作为特定的文本检测来进行处理^[14]。虽然船舶牌照识别与车牌识别的步骤大致相似,但是他们有本质上的区别,与日益成熟的车牌自动识别相比,内河航道船舶名称识别主要难点在于:1)公开数据集的缺少阻碍了船名自动识别的研究;2)航道两岸对船舶的监控角度不尽相同,图像中船名存在倾斜的问题;3)在水路运输系统中,船舶距离岸边较远且船名在船舶上所占比例较小,使得船舶牌照成为较小的目标,对其检测和识别具有一定难度。

为了解决以上问题,本文设计了 ASNI 框架,其中基础框架由 DB^[15]、CRNN^[16]组成。具体地,首先解决数据集缺少的问题,本文数据通过“宜昌区域智慧预警平台”获取,以此制作 SLP 数据集并用于后续的训练和测试。

为了保证视野足够开阔,内河航道两岸的监控大多在具有一定高度的跨江大桥下或两岸监控架杆上,故获取的图像中船名具有一定的倾斜角度,针对此问题,在所提出的 ASNI 框架中引入基于透视变换的倾斜校正模块(corrective module, CT)对船名不规则文本进行矫正。与此同时,针对航道中船舶距离两岸较远且船名在船舶中所占比例较小导致难以识别的问题,本文设计超分辨率重建模块(super-resolution reconstruction module, SR)用于提高图像清晰度以更好的获取图像细节部分,从而更好地识别和分析图像中内容。本文贡献如下:

1) 创建一个包含多种角度、字体和不同时间段的自然场景下内河航道船名数据集 SLP。

2) 设计了一个船名识别框架 ASNI,对内河航道船舶名称有较好的识别效果。

3) 在船名识别部分引入基于透视变换的船名矫正模块和设计超分辨率重建模块,对不规则船名文本进行矫正及提高船舶图像分辨率,以提高在识别方面的性能。

1 相关工作

船舶名称识别可分为船舶名称检测和识别两个子任务,其本质是自然环境下的文本检测和识别,以下是对近几年的文本检测及识别任务和船舶名称检测识别相关研究进行讨论。Tian 等^[17]基于 Faster RCNN^[18]提出 CTPN(connectionist text proposal network)文本检测算法,能对自然场景中竖直方向的文本有较好的检测效果,但是在倾斜文本的检测上表现不佳。Liao 等^[19]提出 TextBoxes 文本检测算法,能够检测多方向和任意形状的文本,但是在复杂背景下容易受到噪声的干扰。近几年,Liao 等^[15]提出的 DB 是较为成熟且在相关领域使用较多的算法,其提出了一种有效的自适应尺度融合模块,通过自适应融合不同尺度的特征来提高尺度鲁棒性,并且能对不规则的文本进行检测。Baek 等^[20]提出 CRAFT 算法,通过对字符级别的区域感知来进行文本检测,能够处理各种不规则形状的文本,但是对于密集文本和小字体表现较差。Shi 等^[16]提出 CRNN 算法,其最大贡献在于能够端到端对不定长的文本序列进行识别。CRNN 是最常用的文本识别模型,尽管其性能略差于最先进的算法,但由于其快速且轻量级,其更具实用性^[21]。

当前已有部分研究结合图像处理方法对船舶进行检

测、分类和识别。金雪丹等^[22]将计算机视觉技术引入船舶交通管理系统,以实现岸台对目标船只的识别,并系统研究了运用图像处理和神经网络技术进行船舶分类和船舶类型识别。古辉等^[23]通过采用一种基于 Hough 变换和 K 均值聚类直线拟合算法矫正变形的牌照图像,并采用快速模板匹配结合垂直与水平投影法对船铭牌上的单个字符进行分割。但他们的识别方法都经过了一系列图像预处理操作,在实时性上不如深度学习方法。相反,几种应用深度学习的船舶名称识别方法为该领域带来了深刻的影响。例如,吴书楷等^[24]提出了一种基于特征融合策略和迁移学习的船名标识字符定位方法,该两阶段船名定位方法能有效减少船名检测误差,但其未考虑图像中存在不规则船名的检测问题。周怡等^[25]提出了基于 YOLO v3 算法的简易目标跟踪方式,实现了对判定为超载的船舶在视频监控上的实时跟踪并识别,但其设计的

船名矫正方法局限于水平方向。Liu 等^[26]提出一种级联中文船舶识别方法,提高区分前景背景和相邻文本行预测方面的性能并消除了密集文本预测的粘性问题,但其未考虑低分辨率图像的识别问题。

尽管一些方法在公开的文本识别数据集或自制的船舶名称识别数据集上取得了很好的效果,但是本文所提出不规则船名识别和低分辨率图像船名识别的问题并没有得到很好的解决。故本文设计一种二阶段船舶识别框架,具体方法将在下节介绍。

2 船舶名称识别方法

本文设计的船舶名称识别框架包括图像采集、图像数据输入、船舶名称定位、船舶识别 4 个步骤完成,流程如图 1 所示。

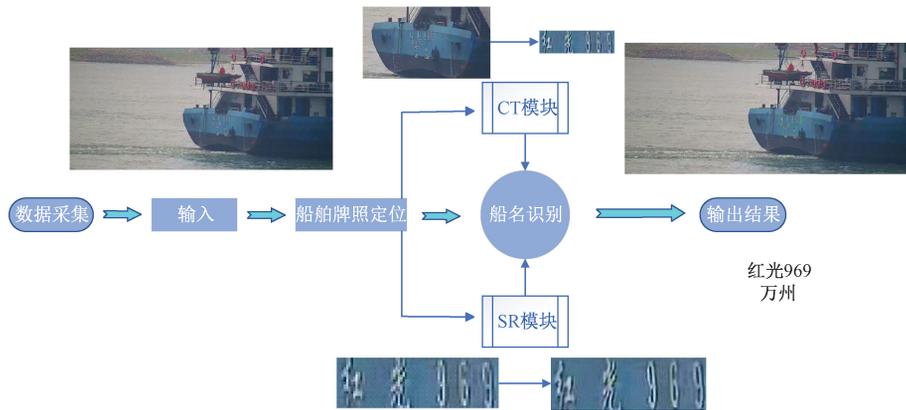


图 1 总体框架流程

Fig. 1 Overall framework process

所提出的 ASNI 包含两个独立组件,即用于船名检测的 DB 和识别的 CRNN。针对 CRNN 不能有效的识别不规则文本和小目标字体^[27],参考基于透视变换的倾斜文本矫正方法对不规则的船名文本进行矫正,保证了后续船名文本的可识别性。同时设计超分辨率重建模块,对分辨率较小的图像进行重建。通过在自制数据集上训练,最终获得一个较高准确率的船舶识别方法,对不规则的船名文本和分辨率较小的船名有较好的识别效果。

2.1 船舶名称检测

对船名进行识别首先需要检测船名位置。如上文所提,可以将船舶名称的定位问题作为特定的文本检测来进行处理^[14]。具体步骤如下:首先,将原始图像输入到特征金字塔主干中;其次,金字塔特征被上采样到相同的尺度并级联以产生特征 F。然后,特征 F 被用于预测概率图和阈值图之后,通过概率图和阈值图计算近似二值映射。最后,在推理阶段,通过检测框生成公式可以从近似二值图和概率图中获得船名检测框,并输出船名样本

图像,船名检测原理如图 2 所示。

概率图标签生成受到 Wang 等^[28]的启发,给定图像中的船名区域的多边形可描述为:

$$G = \{S_k\}_{k=1}^n \quad (1)$$

其中, n 是顶点的数量,然后通过 Vatti 裁剪算法将多边形 G 缩小为 G_s ,从原始多边形周长 L 和面积 A 计算收缩偏移量 D 的计算方式为:

$$D = \frac{A(1-r^2)}{L} \quad (2)$$

其中, r 表示收缩比例,阈值图标签生成方式与其类似。首先使用相同的偏移量 D ,对文本多边形 G 进行膨胀得到 G_d ,然后将 G_s 和 G_d 之间的间隙作为文本区域的边界,通过计算到 G 中最近线段的距离来生成阈值图的标签。

损失函数 L 可以表示为概率映射损失 L_s 、二值映射损失 L_b 和阈值映射损失 L_t 的加权和:

$$L = L_s + \alpha \times L_b + \beta \times L_t \quad (3)$$

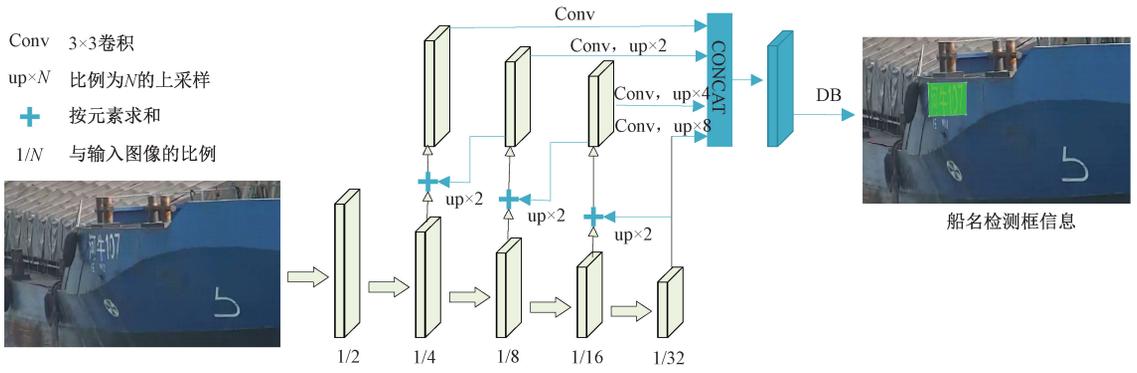


图 2 船名检测原理

Fig. 2 Principle of ship name detection

二值映射用于生成最终的二值化图像,损失设置较大,会使模型更关注对目标的精确定位。设置较大的阈值映射损失权重,使模型更加关注在船名检测中阈值调整的效果,有助于提高模型在不同背景下的适应能力。同时,由于二值映射的结果受阈值映射影响,故 α 和 β 分别设置为 1.0 和 10。其中,对 L_s 和 L_b 应用了二值交叉熵 (BCE) 损失函数:

$$L_s = L_b = \sum_{i \in S_l} y_i \log x_i + (1 - y_i) \log(1 - y_i) \quad (4)$$

S_l 设置为正负比 1 : 3, L_d 则是拓展文本多边形 G_d 内的预测与标签之间的距离之和:

$$L_d = \sum_{i \in R_d} |y_i^* - x_i^*| \quad (5)$$

其中 R_d 是 G_d 内的像素索引集合, y^* 是阈值映射的标签。

在推理阶段,使用概率图或近似二值图来生成船名检测框,他们会产生几乎相同的结果。为了提高效率,使用概率图以便去除阈值分支。检测框的形成过程包括 3 个步骤:用恒定阈值 (0.2) 对概率图或近似二值图进行二值化,得到二值图;从二值映射中获得连接区域 (收缩文本区域);缩小的区域使用瓦蒂剪裁算法以偏移 D' 进行扩张。通过式 (6) 可以反推船名文本区域,具体计算如下:

$$D' = \frac{A' - r'}{L'} \quad (6)$$

其中, A' 是收缩多边形的面积, L' 是收缩多边形的周长,收缩率 r' 根据先验经验设置为 1.5。

2.2 船舶名称识别

船名检测得到待定样本后,经过矫正模块对不规则船名文本矫正和超分辨率重建模块对图像重建,再通过卷积层,对图像提取特征序列,得到特征图。之后循环层对特征序列的每一帧进行预测,对序列中的特征向量学习,并输出预测标签分布。最终在转录层,使用 CTC 损

失把从递归层获取的一系列标签分布转换成标签序列,最终得到船名文本信息。其中,循环层设置为双向 LSTM,用于处理序列信息,对船名文本序列数据中的长距离依赖关系有好的建模能力,能捕捉上下文关系,选择双向 LSTM 为隐层可以更好地满足船名识别任务中对上下文关系、时序信息的需求^[16]。

1) 船名矫正

图像的几何变换方法包括:线性变换、仿射变换和透视变换。但是线性变换和仿射变换都不适用于本方法,他们本质是对图像进行平移或旋转等操作,这些变换都是属于平面变换,对图像中不规则文本矫正效果不佳。透视变换是一种三维空间的投影变换,其在 3 种变换中是最为灵活的一种,考虑到文本矫正需要注意尺寸和位置变化,最终选择透视变换对船名文本进行矫正。透视变换原理如图 3 所示,相关计算如下:

$$\begin{bmatrix} x' & y' & w' \end{bmatrix} = \begin{bmatrix} x & y & w \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \quad (7)$$

$$u = x'/w' \quad (8)$$

$$v = y'/w' \quad (9)$$

其中, $\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$ 为变换矩阵。令 $w = 1$, 得到

公式:

$$u = \frac{a_{11}x + a_{21}y + a_{31}}{a_{13}x + a_{23}y + a_{33}} = \frac{k_{11}x + k_{21}y + k_{31}}{k_{13}x + k_{23}y + 1} \quad (10)$$

$$v = \frac{a_{12}x + a_{22}y + a_{32}}{a_{13}x + a_{23}y + a_{33}} = \frac{k_{12}x + k_{22}y + k_{32}}{k_{13}x + k_{23}y + 1} \quad (11)$$

其中, $k_{mn} = a_{mn}/a_{33}$, ($m = 1, 2, 3$ $n = 1, 2, 3$), (x, y) 为原始图像像素点坐标,经过变换后得到对应的坐标为 (u, v) ^[29]。

2) 超分辨率重建

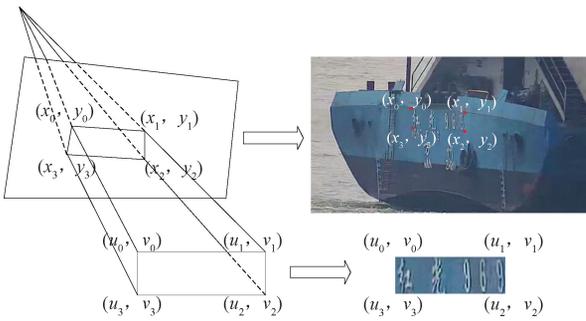


图 3 透视变换矫正原理

Fig. 3 Perspective transformation correction principle

在船舶识别应用场景中,经常面临着仅能获取低分辨率图像的问题,这限制了图像信息可用性和船舶识别的准确性,因此设计超分辨率模块提高待识别图像分辨率至关重要。近些年超分辨率研究使用更快、更深的卷积神经网络在准确性和速度方面取得了突破,但是核心问题却没有得到解决:当超分辨率处理放大倍数时,如何恢复更精细的纹理细节。Ledig 等^[24]结合偏离均方误差 (mean squared error, MSE) 和特征映射定义一种新的感知损失。目标是在不破坏图像信息前提下将低分辨率图像 (low-resolution, LR) 输出成高分辨率图像 (high-resolution, HR)。图 4 显示了超分辨率模块处理图像的部分示例。

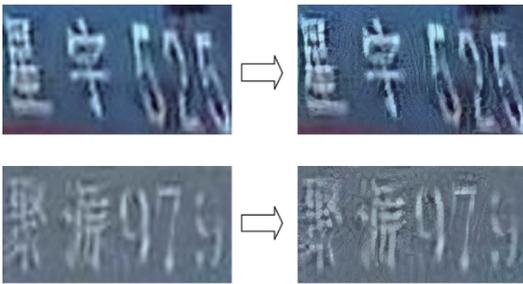


图 4 SR 模块处理效果

Fig. 4 Super-Resolution module processing results

首先通过 LR 图像和对应的 HR 图像训练生成器网络,在训练中使用下采样因子 r 将 I^{HR} 应用于高斯滤波器获取 I^{LR} (I^{LR} , I^{HR} 分别代表 LR 图像和 HR 图像),对于具有 C 个颜色通道的图像,使用大小为 $W \times H \times C$ 的实值张量描述 I^{LR} , $rW \times rH \times C$ 描述 I^{HR} 。将生成器网络训练为由 θ_c (θ_c 表示深度网络的权重和偏差) 参数化的前馈 CNN G_{θ_c} ,其中 $\theta_c = \{W_1:L; b_1:L\}$ 表示 L 层深度网络的权重和偏差,并通过优化特定于 SR 的损失函数 l^{SR} ,最终通过生成器网络训练的生成函数 G 给定输入 LR 图像估计 HR 图像,其中 $n = 1, \dots, N$,相关计算如式 (12) 所示。超分辨率重建模块如图 5 所示,由具有相同布局的 B 个残差块构成,使用两个 3×3 卷积核 (k) 和特征图通道数为 $64(n)$ 的卷积层,其中步长 (s) 设置为 1,最后是批量归一化层^[31] 和使用 ParametricReLU^[32] 作为激活函数。相比于传统的 ReLU,该激活函数引入了一个可学习参数,允许网络根据数据不同部分进行非线性建模,能更好地捕捉船舶名称数据的复杂模式和特征,有助于提高图像的重建质量。

$$\hat{\theta}_c = \underset{\theta_c}{\operatorname{argmin}} \frac{1}{N} \sum_{n=1}^N l^{SR}(G_{\theta_c}(I_n^{LR}), I_n^{HR}) \quad (12)$$

其中,像素级的 MSE 损失计算如下:

$$l_{MSE}^{SR} = \frac{1}{r^2WH} \sum_{x=1}^{rW} \sum_{y=1}^{rH} (I_{x,y}^{HR} - G_{\theta_c}(I^{LR})_{x,y})^2 \quad (13)$$

这也是许多研究所依赖的基于 SR 图像优化目标方法^[33-34]。然而,MSE 优化问题的解决方案往往缺乏高频内容,这导致纹理过于平滑的解决方案在感知上效果不佳。因此,使用更接近感知相似性的损失函数而不依赖像素损失。基于预训练 ResNet 网络的 ReLU 激活层来定义 ResNet 损失。利用 $\phi_{i,j}$,在 ResNet 网络内的第 i 个最大池化层之前通过第 j 次卷积获得特征图。然后,将 ResNet 损失定义为重建图像的特征表示 $G_{\theta_c}(I^{LR})$ 和参考图像 I^{HR} 之间的欧氏距离。内容损失函数如下:

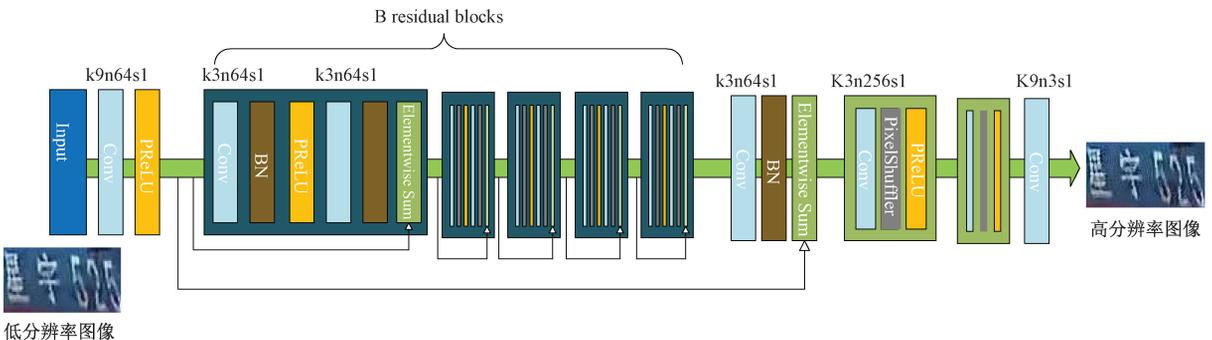


图 5 超分辨率重建模块

Fig. 5 Super-resolution reconstruction model

$$l_{ResNet/i,j}^{SR} = \frac{1}{W_{i,j} H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\phi_{i,j}(I^{HR})_{x,y} - \phi_{i,j}(G_{\theta_c}(I^{LR}))_{x,y})^2 \quad (14)$$

其中, $W_{i,j}$ 和 $H_{i,j}$ 描述 ResNet 网络内各个特征图的维度。

3) 文本识别器

与基于大量数据训练出来的文本矫正算法^[27]不同, 本文提出的船舶识别方法集成了简单的文本矫正方法、超分辨率重建和 CRNN 算法, 在提高多尺度适应性和鲁棒性的同时, 也保证了其推理速度。

文本识别详细方法如下。设船舶检测提取的 ROI 为 $I \in R^{N \times C \times H \times W}$, N 表示感兴趣区域的个数, C 、 H 和 W 分别是提取图像的通道、高度和宽度。为了将 CRNN 作为文本识别器, 将 $I_i \in R^{C \times H \times W}$, 其中 $i = 1, \dots, N$, 后续的 V 表示相应的文本序列集合。得到的样本船名多数长宽比较大, 维持纵横比的前提下, 在训练阶段将图像灰度化并统一缩放至 $32 \times 100 (H \times W)$, 变换后的图像表示为 $I'_i \in R^{C \times H \times W}$ 。 I'_i 通过 CNN 特征提取被编码为特征 $M \in R^{C \times H \times W}$, 其中卷积特征通道 $C = 512, H = 1, W = 25$ 。根据船名文本主要是长序列的特点, 在 CRNN 中下采样设置高为 32, 宽为 4。特征 M 通过映射到序列的操作转换为 $S = [s_1, s_2, \dots, s_n]$, 即将 M 变换成维数为 1×512 的特征向量 w 。在特征序列 S 的顶部上应用 Bi-LSTM^[35] 的隐藏单元, 并将 Bi-LSTM 的输出送到具有 φ 的线性层中, 其中 φ 为转录词典。最后 CTC^[36] 转录层用于序列解码以获得文本结果, CRNN 的优化目标函数为:

$$L = - \sum_{I'_i, V_i} \log p(I_i | y_i) \quad (15)$$

其中, I'_i 和 V_i 是训练图像和相应的标签序列, y_i 是由 I'_i 的递归层和卷积层产生的序列。

3 实验与结果分析

3.1 数据集介绍

本节主要介绍 SLP 的来源及制作规范。由于中文常用字符超过 6 000 个^[37], 在船舶命名中均有可能出现, 这使得船舶的识别任务具有一定的挑战性, 因此数据集的制作至关重要。数据集来源于两个部分, 其中一部分通过智慧预警平台架设在航道两岸的高清摄像头对过往船只进行抓拍获取图像, 另一部分船舶图像数据来源于网络。数据集总共包含 1 135 张停靠在码头或航行中的自然场景下船舶名称的图像, 其中训练集为 908 张, 测试集为 227 张, 包括特写、倾斜、遮挡和不同光照环境等, 部分船名样本如图 6 所示。

对初始数据进行了以下处理: 统一原始数据分辨率, 均为 $1\ 920 \times 900$; 统一命名格式; 使用半自动标注工具



图6 部分船名样本示例

Fig. 6 Partial sample ship names

PPOCRLabel 注释船舶名称, 并确定船舶名称 4 点坐标。为了迎合视角扭曲的船舶名称文本实例, 在注释中使用四边形裁剪出包含船名的预测框。具体地, 以四边形左上角顶点为起点, 通过顺时针旋转生成顶点坐标形式的实例位置标注。最终生成包含船舶名称及位置坐标的标签信息文件, 标签信息用于后续训练和测试。

3.2 实验设置及评估标准

实验流程如下: 将包含船舶名称的图像输入框架后, 首先对船舶名称定位, 将定位的船名裁剪成待定船名样本; 其次, 待定船名样本通过 CT 模块和 SR 模块处理, 输出为待识别船名样本; 最终将船名样本输入文本识别器进行识别并输出船名信息。

实验细节如下: 本文基础模型由百度飞桨提供的预训练模型并在自制数据集上进行微调, 初始学习率设置为 10^{-6} , 训练轮次为 500 轮, 正则化系数为 10^{-5} , 单卡训练时每批大小为 64 个样本。

实验评估指标 FPS 表示每秒处理的图片数量, 精度 (P)、召回率 (R) 和 F-measure (F) 计算如下:

$$P = \frac{TP}{TP + FP} \quad (16)$$

$$R = \frac{TP}{TP + FN} \quad (17)$$

$$F = 2 \times \frac{P \times R}{P + R} \quad (18)$$

其中 TP、FP 和 FN 分别是识别正确的船名文本、识别错误的船名文本和遗漏的船名文本数量。实验环境如表 1 所示:

表1 实验环境

Table 1 Experimental environment

名称	相关配置
CPU	Intel(R) Xeon(R) Silver 4310 CPU @ 2.10 GHz
GPU	NVIDIA GeForce RTX 3090
GPU 显存	24 GB
GPU 加速库	CUDA 11.6.134 CUDNN 8.0
实验平台	Ubuntu 20.04
软件环境	Python 3.7 PaddlePaddle 2.4.1

3.3 消融实验

为了进一步验证本文所设计模块的有效性, 在所述数据集上进行消融实验。为保证消融实验公平性, 对加

入不同模块的识别网络均采用相同的训练策略,详细实验结果如表 2 所示。其中‘√’和‘×’表示是否在框架中应用与之对应的模块。

表 2 模块消融的识别结果

Table 2 Recognition results for modules ablation

方法	CT	SR	P/%	R/%	F/%	FPS
CRNN	×	×	85.37	77.30	81.13	11.79
CRNN	√	×	86.53	79.12	82.65	11.35
CRNN	×	√	86.16	79.84	82.87	10.76
CRNN	√	√	88.69	81.58	84.98	10.14

数据显示,识别模型在应用 CT 模块时有 1.16% 的精度提升,召回率和 F 值分别提升 1.82% 和 1.52%,推理速度略微下降。应用 SR 模块时带来了 0.79% 的精度提升,召回率和 F 值分别提升 2.54% 和 1.74%,推理速度相对于应用 CT 模块下降的较多。应用 CT 模块时对识别精度的贡献相对较大,说明不规则的文本对船名识别的影响较大。加入 SR 模块后对模型的推理速度影响较大,

图像分辨率可能是影响模型推理性能的重要因素之一。通过数据对比可知,在基础框架中应用单个模块时,除 FPS 外均能产生正向优化。由于船舶航行速度有限,ASNI 对推理速度的要求可以适当降低,本框架所提方法导致推理速度略微下降是在可接受范围内,能够满足实际应用的时间需求。

此外,提出的改进方法与基础框架相比,精度和召回率均有所提升,并且每个模块在船名识别中都发挥着重要作用,表明本文所提方法是有效的并具有高准确率的优点。

图 7 显示了船名识别训练中性能增长过程,其中,Accuracy 描述了测试集船名文本识别的平均准确率,Epoch 表示训练轮数。本文框架同时应用两个模块时,达到了 87.50% 的平均准确率,取得了所有方法中最好的识别效果,为现实场景下的船舶名称识别应用提供了可行的解决方案。

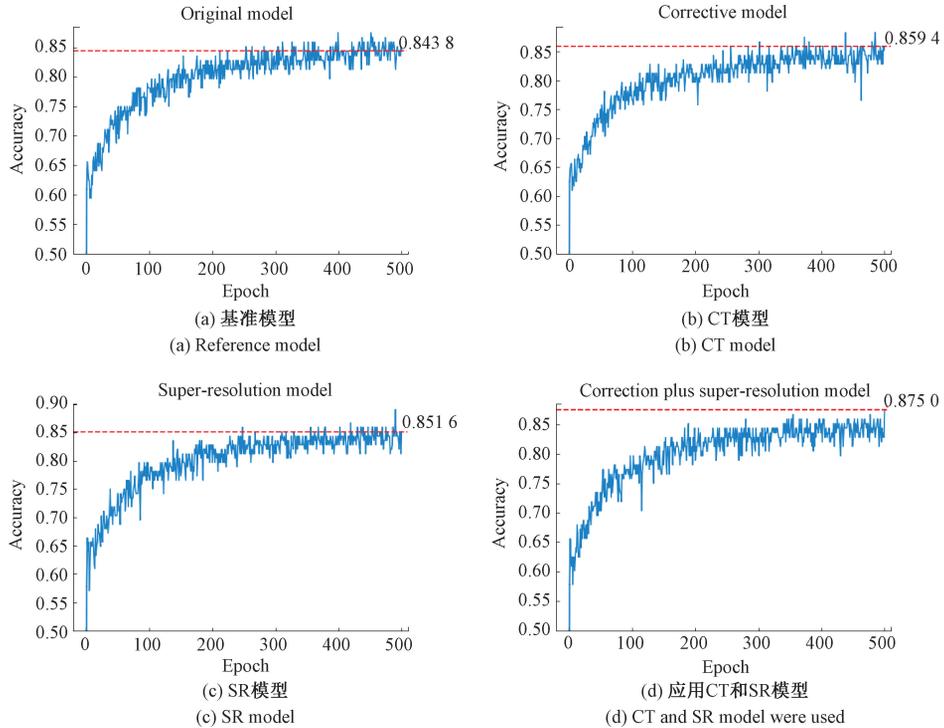


图 7 不同模块的添加对准确率的影响

Fig. 7 Impact of adding different modules on accuracy

3.4 对比实验

为了评估提出方法在船舶名称识别任务上的性能,在 NVIDIA GeForce RTX 3090 上与近几年的相关方法进行了比较。为保证实验公平性,除不同方法的网络结构不同外,训练及测试相关的参数和数据集等均保持一致,相关实验数据如表 3 所示。其中 Model Size 为推理模型大小(MB),FPS 为推理模型每秒处理的图像数量。

表 3 不同方法的准确率对比

Table 3 Algorithm cost and accuracy comparison

方法	Accuracy	Model Size	FPS
本文	87.50	34.9	10.14
ASTER ^[27]	85.39	38.5	9.34
SRN ^[35]	79.25	91.6	5.81
SVTR ^[37]	83.71	45.9	12.05

从表3数据可以看出,本文提出的方法在准确率、模型大小及推理速度上都具有一定的优势。其中SRN方法使用了注意力机制,导致其模型参数过大同时推理速度较慢;SVTR使用单视觉模型,保证了其推理速度达到了最优。本文提出的方法在保持模型最小的同时,也保证了最高的平均准确率,虽然推理速度仅次于最快的SVTR,但是如上一节所提,其完全能够满足船名识别任务的时间需求。同时,给出了不同方法的部分船名识别效果示例,如图8所示,其中识别错误的用红字标出。(由于原始图像不便观看,故图中示例图是经过原始图像裁剪)



图8 不同方法识别结果

Fig. 8 Recognition results of different methods

4 结论

本文基于文本图像修正设计了一种新的级联框架ASNI用于内河航道船舶名称识别。对该框架识别部分进行了有针对性的改进,所提方法有效解决了因分辨率不足和船名文本不规则导致识别不准确的问题,并在多个实验中显示了明显的性能提升,与基础框架相比,该方法在精度、召回率和F-measure上分别提高了3.32%、3.28%和3.85%。同时,通过对比实验验证了本文提出方法的优越性。所提方法不仅有助于提高水路运输系统的安全性,还能为未来进一步探索船名识别技术提供基础。后续的工作将集中在对数据的收集,完善其分类及保证数据多样性,提升框架健壮性使其能应用在更加复杂的环境中。(实验相关代码将开源在:penterlu/Ship-name-recognition (github.com))

参考文献

[1] 宋兵兵. 船舶自动识别系统(AIS)存在的问题及治理路径[J]. 世界海运, 2021, 44(2): 34-37.
SONG B B. Problems and solutions of automatic identification system [J]. World Shipping, 2021, 44(2): 34-37.

[2] YU D, LI X, ZHANG C, et al. Towards accurate scene text recognition with semantic reasoning networks [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 12113-12122.

[3] FANG S, XIE H, WANG Y, et al. Read like humans: Autonomous, bidirectional and iterative language modeling for scene text recognition [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 7098-7107.

[4] DU Y, CHEN Z, JIA C, et al. Svtr: Scene text recognition with a single visual model [J]. arXiv preprint arXiv:2205.00159, 2022.

[5] XU H, ZHOU X D, LI Z, et al. EILPR: Toward end-to-end irregular license plate recognition based on automatic perspective alignment [J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 23(3): 2586-2595.

[6] 丁进超, 张伟伟, 吴训成. 基于双向长短期记忆网络的车牌识别算法 [J]. 电子测量与仪器学报, 2018, 32(6): 173-179.
DING J CH, ZHANG W W, WU X CH. License plate recognition algorithm based on bidirectional long short term memory network [J]. Journal of Electronic Measurement and Instrumentation, 2018, 32(6): 173-179.

[7] ZHANG C, WANG Q, LI X. V-LPDR: Towards a unified framework for license plate detection, tracking, and recognition in real-world traffic videos [J]. Neurocomputing, 2021, 449: 189-206.

[8] 苏云涛, 余红英, 迟进梓. 基于LSTM算法的车牌识别系统方法研究 [J]. 电子测量技术, 2021, 44(18): 67-71.
SUN Y T, YU H Y, CHI J Z. Research on vehicle license plate recognition system method based on LSTM algorithm [J]. Electronic Measurement Technology, 2021, 44(18): 67-71.

[9] ZHANG L, WANG P, LI H, et al. A robust attentional framework for license plate recognition in the wild [J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 22(11): 6967-6976.

[10] SILVA S M, JUNG C R. License plate detection and recognition in unconstrained scenarios [C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 580-596.

[11] ZHANG W, SUN H, ZHOU J, et al. Fully convolutional network based ship plate recognition [C]. 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2018: 1803-1808.

[12] LIU B, WU S, ZHANG S, et al. Ship license numbers recognition using deep neural networks [C]. Journal of Physics: Conference Series, 2018: 012064.

[13] LIU B, LYU X, LI C, et al. Using transferred deep

- model in combination with prior features to localize multi-style ship license numbers in nature scenes [C]. 2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI), 2017;506-510.
- [14] LIU B, SHENG J, DUN J, et al. Locating various ship license numbers in the wild: An effective approach [J]. IEEE Intelligent Transportation Systems Magazine, 2017, 9(4):102-117.
- [15] LIAO M, ZOU Z, WAN Z, et al. Real-time scene text detection with differentiable binarization and adaptive scale fusion [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(1):919-931.
- [16] SHI B, BAI X, YAO C. An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 39(11):2298-2304.
- [17] TIAN Z, HUANG W, HE T, et al. Detecting text in natural image with connectionist text proposal network [C]. Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VIII 14, 2016;56-72.
- [18] GIRSHICK R. Fast R-CNN [C]. Proceedings of the IEEE International Conference on Computer Vision, 2015;1440-1448.
- [19] LIAO M, SHI B, BAI X, et al. Textboxes: A fast text detector with a single deep neural network [C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2017.
- [20] BAEK Y, LEE B, HAN D, et al. Character region awareness for text detection [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019;9365-9374.
- [21] BAEK J, MATSUI Y, AIZAWA K. What if we only use real datasets for scene text recognition? Toward scene text recognition with fewer labels [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021;3113-3122.
- [22] 金雪丹,施朝健. 图像处理与神经网络识别技术在船舶分类中的应用(英文) [J]. 上海海事大学学报, 2007(1):11-16.
- JIN X D, SHI CH J. Application of image processing and neural network recognition technology in ship classification [J]. Journal of Shanghai Maritime University, 2007(1):11-16.
- [23] 古辉,王益义. 一种基于模板匹配的船铭牌字符分割方法 [J]. 浙江工业大学学报, 2010, 38(1):33-35.
- GU H, WANG Y Y. A method for character segmentation of ship nameplate based on template matching [J]. Journal of Zhejiang University of Technology, 2010, 38(1):33-35.
- [24] 吴书楷,刘宝龙,徐舒畅,等. 结合深度特征迁移与融合的两阶段船牌定位算法 [J]. 计算机辅助设计与图形学学报, 2020, 32(4):628-634.
- WU SH K, LIU B L, XU SH CH, et al. Two-stage ship plate location algorithm combining depth feature transfer and fusion [J]. Journal of Computer-Aided Design & Graphics, 2020, 32(4):628-634.
- [25] 周怡,祝啟瑞,谢海成,等. 基于目标检测与模糊匹配的非标船牌识别研究 [J]. 激光与红外, 2021, 51(11):1526-1530.
- ZHOU Y, ZHU Q R, XIE H CH, et al. Research on non-standard ship license recognition based on object detection and fuzzy matching [J]. Laser & Infrared, 2021, 51(11):1526-1530.
- [26] LIU D, CAO J, WANG T, et al. SLPR: A deep learning based Chinese ship license plate recognition framework [J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(12):23831-23843.
- [27] SHI B, YANG M, WANG X, et al. Aster: An attentional scene text recognizer with flexible rectification [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 41(9):2035-2048.
- [28] WANG W, XIE E, LI X, et al. Shape robust text detection with progressive scale expansion network [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019;9336-9345.
- [29] 徐光柱,匡婉,万秋波,等. 基于级联 CNNs 的非约束车牌精确定位 [J]. 计算机工程与科学, 2022, 44(9):1665-1675.
- XU G ZH, KUANG W, WAN Q B, et al. Accurate location of unconstrained license plate based on Cascaded CNNs [J]. Computer Engineering & Science, 2022, 44(9):1665-1675.
- [30] LEDIG C, THEIS L, HUSZÁR F, et al. Photo-realistic single image super-resolution using a generative adversarial network [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017;4681-4690.
- [31] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift [C]. International Conference on Machine Learning, 2015;448-456.
- [32] HE K, ZHANG X, REN S, et al. Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification [C]. Proceedings of the IEEE

International Conference on Computer Vision, 2015: 1026-1034.

- [33] CHARAN K, SHASHANK T, GURURAJ C. Image super-resolution using convolutional neural network [C]. 2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon), 2022:1-7.
- [34] SHI W, CABALLERO J, HUSZÁR F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016:1874-1883.
- [35] GRAVES A, MOHAMED A R, HINTON G. Speech recognition with deep recurrent neural networks [C]. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 2013:6645-6649.
- [36] GRAVES A, FERNÁNDEZ S, GOMEZ F, et al. Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks [C]. Proceedings of the 23rd International Conference on Machine Learning, 2006:369-376.
- [37] LIU Y, SHEN C, JIN L, et al. Abcnet v2: Adaptive bezier-curve network for real-time end-to-end text spotting [J].

IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(11):8048-8064.

作者简介



卢鹏涛, 2021年于三峡大学获得学士学位, 现为三峡大学硕士研究生, 主要研究方向为计算机视觉、图像处理和深度学习。

Lu Pengtao received his B. Sc. degree from China Three Gorges University in 2021.

He is currently a M. Sc. candidate at China Three Gorges University. His main research interests include computer vision, image processing and deep learning.



汪方毅 (通信作者), 2017年于华南理工大学获得博士学位, 现为三峡大学讲师, 主要研究方向为多模态信息处理和脑影像分析。

E-mail: fy_wang@ctgu.edu.cn

Wang Fangyi (Corresponding author) received his Ph. D. from South China University of Technology in 2017. He is currently a lecturer at China Three Gorges University. His main research interests include multimodal information processing and brain image analysis.